Bilingualism: Language and Cognition

cambridge.org/bil

Contributions of declarative and procedural memory to accuracy and automatization during second language practice

Diana Pili-Moss[1], Katherine A. Brill-Schuetz[2], Mandy Faretta-Stutenberg[3] and Kara Morgan-Short[4]

[1]Department of Linguistics and English Language, University of Lancaster; [2]Department of Psychology, University of Illinois at Chicago; [3]Department of World Languages and Cultures, Northern Illinois University and [4]Department of Hispanic and Italian Studies, and Department of Psychology, University of Illinois at Chicago

Research Article

Cite this article: Pili-Moss D, Brill-Schuetz KA, Faretta-Stutenberg M, Morgan-Short K (2020). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition* **23**, 639–651. https://doi.org/10.1017/S1366728919000543

Received: 24 July 2018
Revised: 4 July 2019
Accepted: 16 July 2019
First published online: 1 October 2019

Keywords:
declarative memory; procedural memory; L2 individual differences; L2 practice; L2 automatization

Address for correspondence:
Diana Pili-Moss,
E-mail: d.pilimoss@lancaster.ac.uk

Abstract

Extending previous research that has examined the relationship between long-term memory and second language (L2) development with a primary focus on accuracy in L2 outcomes, the current study explores the relationship between declarative and procedural memory and accuracy and automatization during L2 practice. Adult English native speakers had learned an artificial language over two weeks (Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter & Wong, 2014), producing four sessions of practice data that had not been analyzed previously. Mixed-effects models analyses revealed that declarative memory was positively related to accuracy during comprehension practice. No other relationships were evidenced for accuracy. For automatization, measured by the coefficient of variation (Segalowitz, 2010), the model revealed a positive relationship with procedural memory that became stronger over practice for learners with higher declarative memory but weaker for learners with lower declarative memory. These results provide further insight into the role that long-term memory plays during L2 development.

Introduction

Cognitive and psycholinguistic approaches to second language (L2) acquisition in the last twenty years have looked to individual differences as a means to understand the mechanisms that support L2 development and have examined several domain-general factors including, for example, executive function, short-term memory, working memory and, more recently, long-term memory. According to bipartite models of the architecture of long-term memory, DECLARATIVE MEMORY is a system capable of fast learning and retention of information relative to events, facts and arbitrary associations, whereas NONDECLARATIVE MEMORY is a system comprised of several subsystems, one of which is PROCEDURAL MEMORY, which consolidates information more gradually and is largely responsible for implicit sequence learning, probabilistic learning and motor skill learning (e.g., Cabeza & Moscovitch, 2013; Eichenbaum, 2008, 2011; Squire, 2004; Squire & Dede, 2015; Squire & Wixted, 2011).

A number of recent correlational studies in second language acquisition (SLA; e.g., Antoniou, Ettlinger & Wong, 2016; Brill-Schuetz & Morgan-Short, 2014; Ettlinger, Bradlow & Wong, 2014; Hamrick, 2015; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter & Wong, 2014; Morgan-Short, Finger, Grey & Ullman, 2012; Pili-Moss, 2018; Suzuki, 2017; see also Hamrick, Lum & Ullman, 2018 for a recent meta-analysis) have investigated the relationship between L2 learning outcomes and specific memory-dependent declarative and procedural learning abilities, assessed by means of behavioral tasks that have been independently linked to declarative and procedural memory in the neuropsychological literature. Generally, these studies have evidenced a positive relationship between learning outcomes and long-term memory measures, although this may be modulated by a range of factors (e.g., type and amount of input, level of proficiency, linguistic structure, type of instruction).

In addition to understanding the role of declarative and procedural memory on L2 learning outcomes, it is undoubtedly of interest to SLA researchers to gain a more complete picture of how memory modulates L2 learning during practice. However, only two studies to date (Pili-Moss, 2018; Suzuki, 2017) have examined this issue. Extending the analysis of data collected but not discussed in Morgan-Short et al. (2014) and Morgan-Short, Deng, Brill-Schuetz, Faretta-Stutenberg, Wong and Wong (2015), the aim of this paper is to address this gap in the literature and elucidate the role that declarative and procedural learning ability play in modulating accuracy and automatized language processing during practice.

© Cambridge University Press 2019

CAMBRIDGE UNIVERSITY PRESS

Cognitive models of L2 learning

Recent approaches to the organization of memory have informed our theoretical understanding of L2 acquisition. In particular, three cognitive models of late-learned L2 have posited the relevance of declarative and procedural memory (or knowledge) for L2 learning (DeKeyser, 2015; Paradis, 2009; Ullman, 2004, 2015, 2016). According to Ullman's (2004, 2015, 2016) declarative/procedural model (DP model), declarative and procedural memory are largely independent neural memory systems and their activity is modulated by a range of external and internal factors including hormonal and genetic factors, age, and sex. Under certain circumstances, declarative and procedural memory can also interact cooperatively or competitively: for example, in case of functional impairment or attenuation of one of the systems.

In Ullman's model (Hamrick et al., 2018; Ullman, 2004, 2015, 2016) the two systems generally underlie the learning of different types of linguistic knowledge. More specifically for first language (L1), Ullman's model posits that declarative memory primarily supports the learning and use of all aspects related to lexis as well as idiosyncratic forms (e.g., irregular morphology) and 'chunks'. Procedural memory supports the learning and use of (hierarchical) sequences and rules across different linguistic domains (including syntax, morphology and possibly phonology). With regard to L2 acquisition, Ullman's model predicts that declarative memory will support the learning of lexis at all stages of exposure and levels of proficiency. Declarative memory is also expected to support the learning of L2 grammar at early stages of exposure/proficiency. Procedural memory, however, is expected to play an increasingly stronger role for L2 grammar at later stages of exposure, when learners have had more practice with the L2.

Paradis' (2009) model makes similar claims as Ullman's model, but differs from it in at least three respects. First, concerning lexis, Paradis posits that declarative memory is only responsible for the learning of form-meaning relationships (vocabulary), whilst learning of word subcategorization patterns (lexicon) depends on procedural memory. Secondly, Paradis' model assumes that language processing in declarative memory leads to explicit (conscious) representations, whilst, according to Ullman (2015), declarative processing does not necessarily imply consciousness (Henke, 2010). Finally, Paradis (2009) largely limits the role of procedural memory to the L1 and, although it is not excluded, L2 procedural processing is considered to be "very rare in practice" (p.16).

From a slightly different perspective focused on L2 knowledge, DeKeyser (2015) has proposed the Skill Acquisition model with roles for declarative and procedural knowledge in L2 development and automatization. The model distinguishes three phases in the automatization process. In the declarative stage, the learner relies exclusively on declarative knowledge (in the form of explicitly taught or induced linguistic rules). The second stage (proceduralization) is a relatively early phase in practice in which declarative knowledge is "acted on" (DeKeyser, 2015, p. 95), resulting in the creation of increasingly procedural/behavioral representations of the initial knowledge. At this stage, learners increasingly draw on both types of knowledge as language rules are practiced, and they no longer need "to retrieve bits and pieces of information from memory to assemble them" (DeKeyser, 2015, p. 95). Although there is no transfer of information or transformation of knowledge from declarative to procedural, a strong declarative knowledge is argued to support the onset of proceduralization (DeKeyser, 2015). In the last stage (automaticity), language knowledge is fully proceduralized in that its use is both rapid and accurate, although declarative knowledge representations may be maintained.

Because it is specified in regard to type of linguistic knowledge (the product of learning), DeKeyser's model is largely independent from assumptions about the structure of neural memory systems. However, transposing the relationship between declarative and procedural knowledge to the memory systems that encode them, DeKeyser's model would be compatible with the prediction of a substantial involvement of declarative memory in the initial stages of practice, followed by an increasingly stronger reliance on procedural memory as language processing becomes proceduralized and then automatized. Thus, notwithstanding the highlighted differences between Ullman's and Paradis' approaches, as well as the slightly different focus on memory versus knowledge in the different models, perspectives based on the characteristics of neural memory systems and type of L2 knowledge make generally consistent predictions for the role of declarative and procedural memory and knowledge in L2 development and automatization.

Declarative and procedural learning ability as individual differences in L2 development

In a recent meta-analysis Hamrick et al. (2018) found that, for L2 adults, lexical abilities were consistently related to declarative memory, whilst grammatical abilities were related to declarative memory at early stages of exposure (see also Faretta-Stutenberg & Morgan-Short, 2018; Hamrick, 2015; Morgan-Short et al., 2014, Pili-Moss, 2018, Study 2) and to procedural memory at later stages of exposure (see also Brill-Schuetz & Morgan-Short, 2014; Faretta-Stutenberg & Morgan-Short, 2018; Hamrick, 2015; Morgan-Short et al., 2014; Pili-Moss, 2018, Study 1, for a different pattern of results in children).

In one of the studies included in the meta-analysis, Morgan-Short et al. (2014) exposed 14 university students to Brocanto2, a miniature language based on Spanish, under an implicit training condition in which participants were told that they would be learning an artificial language but were not provided with metalinguistic information or direction to search for rules (DeKeyser, 1995; Norris & Ortega, 2000, p. 437). It is important to note that no assumption was made about the type of knowledge acquired by the learners (implicit or explicit). After initial passive, meaningful aural exposure, the participants practiced language comprehension and production in the context of a computer board game (4 sessions over 2 weeks, for a total of 72 game blocks; see Methods section for further details). Two versions of an aural grammaticality judgment test (GJT) were administered respectively at the end of the first session and at the end of practice as the L2 outcome measure. Results on these GJTs showed that declarative learning ability significantly predicted language development after the first session, whilst procedural learning ability was a significant predictor of development at the end of the experiment.

Beside stage of exposure, other studies have provided evidence for additional factors that may modulate the role of long-term memory abilities (for reviews see Buffington & Morgan-Short, 2019; Hamrick et al., 2018). Some of these include order of presentation in the input (Antoniou et al., 2016), type of rule (Antoniou et al., 2016; Ettlinger et al., 2014; Pili-Moss, 2018), type of training condition and learning context (Brill-Schuetz &

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

Morgan-Short, 2014; Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018), processing speed-up (Suzuki, 2017), and age (Pili-Moss, 2018). A further modulating factor that has been recognized in the literature (e.g., Hamrick et al., 2018; Morgan-Short et al., 2014), but has not been directly investigated to date, is the role of type of task. It could be argued that, due to their specific characteristics, tasks may differ in the way they engage declarative or procedural processing. More generally, type of task could also refer to whether the task is an assessment task (e.g., a GJT) or a learning task involving more extended language practice.

To our knowledge, only one study has examined the role of long-term memory during L2 practice. Pili-Moss (2018, Study 2) trained 36 L1 Italian university students in a version of Brocanto2 based on Japanese (BROCANTOJ), using the same board game context and training condition as Morgan-Short et al. (2014). However, in this case the training was shorter (6 blocks over 3 consecutive days, corresponding to the very initial stages of L2 learning), included only comprehension practice, and tracked the effects of declarative and procedural learning ability during practice in addition to administering a GJT at the end of practice. Given the comparatively more limited exposure to the language, the GJT results were consistent with Morgan-Short et al. (2014), indicating that declarative learning ability, but not procedural learning ability, significantly predicted L2 accuracy at early stages of learning. The study also found that declarative learning ability significantly predicted accurate performance during practice, although for a subset of stimuli (sentences for which the comprehension of links between word order and thematic interpretation was crucial), a significant positive interaction between declarative and procedural learning ability was also evidenced, indicating that both cognitive abilities may contribute to the learning of the mapping rules linking thematic roles and syntactic linearization of arguments in a second language.

Overall, with the exception of Pili-Moss (2018), studies that investigated the relationship between L2 development and long-term memory have provided insight into how these individual differences may support L2 learning as assessed by outcome measures taken at one or two discrete points in the learning process. For this reason, it could be argued that they provide only partial insight into the role cognitive variables play in the learning process. Studies offering a more fine-grained measure of the relationship between long-term memory individual differences and L2 development during practice have the potential to provide more direct insight into how this relationship develops over time. Such research may be all the more informative if it considers indices of L2 development beyond accuracy: for example, neurocognitive processing of L2 (Faretta-Stutenberg & Morgan-Short, 2018) or automatization (Suzuki, 2017).

L2 automatization in L2 learning

An important aspect of L2 assessment in SLA research is the study of language automatization, i.e., the extent to which L2 processing in comprehension and production can reach levels of fluency approaching those of L1 speakers in nonnative language users (DeKeyser, 2007; Segalowitz, 2010). Automaticity in language comprehension and production is characterized by processing that is stable, fast, ballistic (i.e., unstoppable once triggered), not controlled and not limited by working memory capacity, and is qualitatively defined in opposition to similar processing that does not present automatic characteristics, i.e., is unstable, slow, controlled, stoppable, possible only within the limits of working memory capacity, etc. (Segalowitz, 2003; 2013).

Measures of reaction time (RT) decrease over time have been used as one of the main indices in the operationalization of automatization (including in L2 linguistic processes). For example, following approaches to skill acquisition developed in the ACT-R framework (e.g., Anderson 1993, 2007), some L2 studies (e.g., DeKeyser, 1997, Ferman, Olshtain, Schechtman & Karni, 2009) have measured the automatized status of L2 processing during practice by assessing the extent to which the reduction of RTs over time can be fitted to a power function.

Other authors (e.g., Segalowitz, 2010; Segalowitz & Segalowitz, 1993) have argued that a measure of automatization should capture the fact that automatized language processing becomes NOT ONLY FASTER BUT ALSO LESS VARIABLE as a function of practice. As an alternative automatization measure they have proposed the coefficient of variation (CV), an index that equals the ratio between the intraindividual standard deviation and the mean RT. When RTs are decreasing, a simultaneous CV decrease is the result of a more than proportional reduction in the standard deviation, indicative of a qualitative restructuring of the process. According to Segalowitz (2010), two minimal conditions should be simultaneously observed for the index to constitute reliable evidence of automatization: (a) a significant decrease of both the CV and the RT over the course of practice (or at different points of testing or in group comparisons), and (b) a significant positive correlation between CV and RT.

SLA studies that have used the CV index have investigated L1/L2 differences in lexical access (e.g., Akamatsu, 2008; Phillips, Segalowitz, O'Brien & Yamasaki, 2004; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz & Wood, 1998; Segalowitz, Trofimovich, Gatbonton & Sokolovskaya, 2008) and, more recently, L2 grammar learning (e.g., Hulstijn, Van Gelderen & Schoonen, 2009; Lim & Godfroid, 2015; Ma, Yu & Zhang, 2017; Suzuki, 2017; Suzuki & Sunada, 2018). In general, CV studies on lexical access have found consistent evidence of automatization, whilst the evidence for L2 grammar learning has been mixed.

For example, Hulstijn et al. (2009, Experiment 1) investigated the development of automatization in 397 L1 Dutch high-school learners of English. The longitudinal study analyzed RT data from four computerized tasks administered to the students in the L1 and the L2 once a year, in Grade 8 (13–14 years of age), 9, and 10. The tasks administered were a word/nonword discrimination task, a lexical retrieval task, a sentence verification task (based on semantic acceptability) and a sentence completion task (probing grammaticality). Overall, the study found only partial evidence of automatization in terms of significant CV decrease and CV/RT correlations, and mainly in the lexical-based tasks. Based on their results the authors questioned the use of the CV as an index of automatization, suggesting that it may be too restrictive. However, as noted in Lim and Godfroid (2015), the length of training per se does not ensure that automaticity will be attained. Arguably, this may be especially the case if practice and testing take place in different environments requiring a TRANSFER of automatized skilled behavior across different conditions/tasks (on this point see also DeKeyser, 2007; Suzuki & Sunada, 2018).

Lim and Godfroid (2015) conceptually replicated Hulstijn et al. (2009) assessing automatization in 40 Korean L2 learners of English (20 intermediate and 20 advanced) and 20 L1 English speakers. The testing included a lexical discrimination task (based on animacy), in addition to a sentence completion

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

task and a sentence plausibility task similar to those deployed in Hulstijn et al.'s original experiment. For the sentence completion task, a cross-sectional comparison of the three groups found significant CV decreases as a function of language proficiency together with significant CV/RT correlations for both intermediate and advanced L2 learners. In a similar study, Ma et al. (2017) compared low and high proficiency Chinese learners of English in a sentence plausibility task and also found a significantly lower CV in high-proficiency learners. Overall, the results of cross-sectional studies seem to suggest significant decreases in the CV index (i.e., an increase in automatization) as a function of proficiency at least for some of the tasks tapping the development of L2 grammar.

To date only Suzuki (2017) investigated the extent to which L2 automatization is modulated by long-term memory (procedural learning ability). Sixty L1 Japanese university students in two experimental groups (short and long spacing) were exposed in explicit instruction conditions to verbs with present progressive morphology in a miniature language across four sessions, 3.3 days or 7 days apart. CV decreases relative to two oral production tests administered at the beginning and at the end of each session did not provide evidence of automatization.

Further, procedural learning ability (measured by the Tower of London task - TOL) was found to significantly correlate with RT decrease in the short-spacing condition, but no significant relationships were found between procedural learning ability and CV. Overall, Suzuki (2017) extended previous research on the relationship between long-term memory abilities and accuracy to speed-up. However, the extent to which these abilities may contribute to automatization remains an open question.

Motivation for the study and research questions

Based on an analysis of practice data that were not reported or analyzed in Morgan-Short et al. (2014) or Morgan-Short et al. (2015), the aim of the present study was to explore the role of declarative and procedural learning ability in L2 development during practice over time in regard to accuracy (in comprehension and production) and automatization (in comprehension). For the current analysis, participant responses on comprehension and production practice trials are used to examine accuracy during practice and CV is calculated based on the reaction times in the comprehension blocks as an index of automatization. As RTs were not available for production blocks, automatization in production is not investigated in the present study. The research questions were formulated as follows:

RQ1: To what extent do declarative and procedural learning ability predict accuracy in comprehension and production during L2 practice? Do these effects differ across various stages of practice?

RQ2: To what extent do declarative and procedural learning ability predict automatization in comprehension during L2 practice? Do these effects differ across various stages of practice?

For RQ1, based on Morgan-Short et al. (2014) and Pili-Moss (2018, Study 2), we hypothesize a significant role of declarative learning ability in supporting L2 accuracy early in practice. Further, if the pattern of effects in the practice data is comparable to the one found in the GJT (Morgan-Short et al., 2014), we also expect an attenuation of the effect of declarative learning ability at late stages of training, possibly accompanied by an increasingly stronger effect of procedural learning ability. For automatization, based on theoretical assumptions in DeKeyser (2015) and Ullman (2015; 2016), we hypothesize (a) that declarative learning ability will have a significant role early in practice, followed by an increase in the effect of procedural learning ability as practice progresses, and (b) that declarative learning ability will act as a facilitating factor in the automatization process supporting the transition from the declarative to the proceduralization stage.

Methods

The current study is an analysis of data collected but not reported or examined by Morgan-Short et al. (2014) and Morgan-Short et al. (2015). In regard to the relationship between long-term memory individual differences data (collected during a cognitive test session) and L2 development, these previous studies examined results based on the L2 outcome measure (the GJT) administered during two L2 assessment sessions. In contrast, the current study examines L2 data collected during the four language training and practice sessions. Below we provide an overview of the participants and of the materials and procedures related to the cognitive test session and the language training and practice sessions. We do not describe the assessment sessions, as these data were not relevant to the current study (for full reports see Morgan-Short et al., 2014; Morgan-Short et al., 2015).

Participants

Data from 14 participants (6 female) were analyzed in the current study. The participants were right-handed, healthy young adults (mean age = 22.21, $SD = 2.72$) who were native speakers of English, spoke 1.21 non-native languages ($SD = 0.58$), and had limited exposure to Romance languages. Six additional participants began the study but were excluded from analysis for various reasons. See Morgan-Short et al. (2014) for more details about the participants, participant attrition, exclusion and compensation.

General procedure

Seven experimental sessions had been scheduled over a two-week period, one to three nights apart. The cognitive tests, including an IQ assessment (Kaufman & Kaufman, 2004), were administered with counterbalanced order across participants in Session 1 (approximately 3 hours). The remaining sessions were devoted to language training and practice (Sessions 2, 4, 5, and 6) and assessment (Sessions 3 and 7) and lasted on average 2.6 hours and 1 hour respectively.

Materials and Procedures

Cognitive tests

Participants completed two measures of declarative and two measures of procedural learning ability and composite scores for each were obtained. Part V of the Modern Language Aptitude Test (MLAT-V; Carroll & Sapon, 1959) was administered as a verbal measure of declarative learning ability. For this task, participants learned 24 pseudo-Kurdish and English word association pairs and subsequently completed a four minute, 24-item, multiple-choice test where they chose the English equivalent for each pseudo-Kurdish word. MLAT-V scores reflect the total number of correct responses. The Continuous Visual Memory Task

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

(CVMT; Trahan & Larrabee, 1988) was administered as a non-verbal measure of declarative learning ability. For this task, participants viewed a series of abstract designs presented on a computer screen for 2 seconds, and indicated whether each design was novel (63 items presented once each) or had appeared previously (7 items presented 7 times interspersed throughout the novel items). Participants' responses were used to calculate a CVMT *d'* score.

The measures of procedural learning ability were a computerized version of the Tower of London task (TOL; Kaller, Unterrainer & Stahl, 2012; Kaller, Rahm, Köstering & Unterrainer, 2011; Unterrainer, Rahm, Leonhart, Ruff & Halsband, 2003) and a dual-task version of the Weather Prediction Task (WPT; Foerde, Knowlton & Poldrack, 2006). In the TOL, participants were asked to click and drag ball-like shapes on pegs, from an initial configuration to a goal configuration, in a specified number of moves (ranging from 3 to 6). Comparing the initial and the final trials for each set, the decrease in the reaction time between the presentation of the initial configuration and the first move (initial think time) was used as the measure of procedural learning ability. In the WPT, participants select a weather prediction ("sunshine" or "rain") based on patterns of four different "tarot cards" presented on the computer (320 trials in 8 pseudorandomized blocks). Each combination of cards, displayed for 3 seconds, represents a different probability for "sunshine" or "rain." After each response, the correct answer is displayed on the screen. The distractor task required participants to count high tones (1000 Hz) presented along with low tones (500 Hz) throughout each block. After excluding trials for which the probability was 50%, accuracy on the final dual-task block was used as the WPT score.

Artificial language

The artificial language, Brocanto2 (Morgan-Short, 2007; Morgan-Short, Sanz, Steinhauer & Ullman, 2010; Morgan-Short, Finger, Grey & Ullman, 2012; Morgan-Short, Steinhauer, Sanz & Ullman, 2012), was modeled after Brocanto (Friederici, Steinhauer & Pfeifer, 2002). Brocanto2 has 13 lexical items: 4 nouns (*pleck, neep, blom, vode*), 2 adjectives (*troise/o, neime/o*), 1 article (*li/u*), 4 verbs (*klin, nim, yab, praz*) and 2 adverbs (*noyka, zayma).* Nouns have gender (masculine or feminine) and agree with adjectives and articles. Brocanto2 has a productive structure consistent with natural languages, can be spoken and understood within a meaningful context and displays the SOV word order as shown in (1).

(1) (Noun-Adjective-Article) - (Noun-Adjective-Article) – Adverb – Verb

Each Brocanto2 sentence describes a move on a computer board game whose rules are completely independent from the rules of the language. In Brocanto2, the nouns represent the four game tokens of the game, and the adjectives describe the tokens' shape (round or square). The four Brocanto2 verbs indicate the game moves: move, swap, capture, and release. The two adverbs indicate whether moves are in the horizontal or vertical direction.

Vocabulary training

At the start of each of the four training and practice sessions, computer-based vocabulary training was administered. The program individually presented Brocanto2 lexical items auditorily, with the matched visual symbols that represented their meanings. Participants trained at their own pace and were tested when they believed that they had learned all the lexical items. During the vocabulary test, each symbol was presented twice and participants were asked to state out loud the lexical item that corresponded to it. If participants did not achieve a score of 100% accuracy on this test, they repeated vocabulary training and took the test again until they reached criterion.

Language training

In each training and practice session, after vocabulary testing, learners were auditorily exposed to 129 Brocanto2 phrases and sentences in association with the visual representation of the corresponding game token or move on the computer game board. The timing of the training was pre-determined (approximately 13.5 minutes), and learners were asked to pay attention as they would take a short quiz about what they saw after the training.

Language practice

Language practice, administered after language training, occurred in the context of the computer-based game. It consisted of 72 alternating comprehension and production modules (36 modules each; 20 novel sentence stimuli per module). During comprehension modules, participants heard sentences in the language and were instructed to "make the move on the game board that corresponds to the statement you heard." For each comprehension trial, accuracy and RTs (measured in milliseconds from the end of the playback of the aural stimulus to the move completion) were recorded by the computer. During production modules, participants saw a move and were instructed to "state the move out loud" by using a Brocanto2 sentence. For each production trial, accuracy was entered into the computer by the researcher. For all comprehension and production trials, the computer provided immediate feedback on whether their response was correct or incorrect. No additional information or opportunity to modify the response was provided. Participants completed 12 practice modules during Session 2 and 20 practice modules in each of the three subsequent training and practice sessions.

Analyses and Results

RQ1

Descriptive statistics

For descriptive statistics purposes, mean block accuracy was calculated for comprehension and production practice across participants (Table 1) for each of the four training and practice sessions. The data show that accuracy was relatively high for comprehension as early as the second session (on average 16.7 accurate responses per block out of 20). By the end of training it had increased on average to 18.6 accurate responses per block out of 20, with a small standard deviation. For production, accuracy developed more slowly over time reaching a maximum average of 18 accurate responses per block out of 20 with higher variability among participants.

For preliminary insights into any relationship between declarative and procedural learning ability and accuracy during practice, correlations were run between mean block accuracy for comprehension and production and declarative and procedural learning ability (Table 2). Declarative learning ability showed medium to large relationships (Plonsky & Oswald, 2014) with accuracy in comprehension throughout training, as well as an overall marginally statistically significant correlation. By contrast, the relationship

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

Table 1. Mean accuracy per block across sessions in language comprehension and production (N = 14).

	S1	S2	S3	S4
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Comprehension	11.3 (3.7)	16.7 (3.1)	18.4 (1.6)	18.6 (0.5)
Production	3.5 (5.8)	11.2 (7.9)	14.9 (6.6)	18.0 (2.9)

Note: Maximum score per block = 20.

Table 2. Correlations between accuracy and declarative and procedural learning abilities across sessions for comprehension and production practice (N = 14).

	S1	S2	S3	S4	Overall
			Comprehension practice		
Declarative	.613	.633	.511	.429	.692^
Procedural	.163	.078	.161	.162	−.038
			Production practice		
Declarative	.652	.388	.351	.364	.475
Procedural	.381	.150	.304	.277	.301

Note: ^p < .10; *p < .05. Bootstrapped; Holm-Bonferroni corrected.

between procedural learning ability and accuracy in comprehension was weak throughout the training. For accuracy in production, small to large relationships were evidenced for declarative learning ability with no statistically significant correlations. Only small relationships were evidenced for procedural learning ability and accuracy in production. Thus, a comparatively stronger role of declarative learning ability in supporting accuracy was found for comprehension but not production. A Pearson's correlation was also run between the declarative and procedural memory scores and showed that the relationship between the two variables was positive but not significant (r = .222; p = .466, bootstrapped).

Data modeling

In order to directly address RQ1, two separate analyses were conducted for comprehension and production accuracy. Data modeling was performed using binomial generalized mixed-effects models (Faraway, 2016) with the *glmer* function (*lme4* package, Bates, Maechler & Bolker, 2011) in the *R* environment (R Development Core Team, 2018). In both accuracy models, the outcome variable was a measure of the log-likelihood that individual comprehension/production trials were correct given a one-unit increase in the predictor variables. The main effects included Session (treated as a continuous and centered variable) and the two main predictors of interest, declarative and procedural learning ability (which were already available as standardized measures in Morgan-Short et al., 2014 and are abbreviated as Decl and Proc, respectively). Interactions were added if they statistically significantly improved the fixed-effects model's fit (as determined by the likelihood ratio test). To determine the structure of random effects, we first ascertained that both random effects of participants and trial items on intercepts improved the fixed-effects model. We fit the maximal random effect structure (Barr, Levy, Scheepers & Tily, 2013) to the extent justified by the data. A random slope was included in the final model if the model converged and the random slope significantly improved the model's fit compared to the next simpler nested model (as determined by the likelihood ratio test). In both models, a positive β coefficient indicated a positive correlation between the predictor and the log-likelihood of a trial being correct, whilst a negative β value indicated a negative correlation between the predictor and the log-likelihood of a trial being correct. The syntax of all final models is reported in the supplementary materials S1 (Supplementary Materials). The interpretation of the models' effect size (R^2) follows the field-specific recommendations in Plonsky and Ghanbar (2018).

Accuracy in comprehension

The model for comprehension (Table 3) was derived after ensuring that the risk of multicollinearity between the predictors was low (condition number = 1.24). Overall, the model accounted for 56% of the variance compared to 26% in the corresponding model where random effects were not included (all effects computed using R^2).

The model yielded a positive, statistically significant effect of Session on accuracy (p < .001), indicating that the log-likelihood that items were produced correctly increased significantly as training progressed (a medium effect; R^2 = .47). Turning to the predictors of interest, the model outcome was that, overall, declarative learning ability was a statistically significant positive predictor of accuracy (p < .001) with a medium effect size (R^2 = .30). By contrast, procedural learning ability had a positive but nonsignificant relationship with accuracy with a negligible effect size (R^2 = .01). The β coefficient of the Decl by Session interaction indicated that the effect of declarative learning ability decreased, although nonsignificantly, across practice. The plot in Figure 1 illustrates the fairly consistent effect of declarative learning ability at three subsequent stages corresponding to intervals representing early, middle, and later stages of practice.

Accuracy in production

After testing multicollinearity (condition number = 1.24), the model of the production data was derived. Overall, the final model (Table 4) explained about 88% of the variance, compared to 43% in the corresponding model where random effects were not specified. Note that this implies that random effects are likely to have had a substantial influence on the initial correlation results (cf., descriptive statistics; Table 2), a fact that would account for the lack of perfect alignment between the results of the initial correlation and the final model's results.

The model returned a positive statistically significant, large effect of Session on accuracy (R^2 = .84, p < .001), indicating that the log-likelihood that items were produced correctly increased significantly as training progressed. Both declarative and procedural learning ability had positive, though nonsignificant, medium-sized effects (R^2 = .36 and R^2 = .45, respectively). The Proc by Session interaction was found to be statistically significant (p < .001), and its negative β coefficient indicated a significant decrease in the ability of procedural learning ability to predict

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

Table 3. Mixed-effects model of the effects of session, declarative learning ability and procedural learning ability on accuracy in comprehension.

				95% CI		
Fixed effects	β	*SE*	*z*	lower	upper	*p*
(Intercept)	2.68	0.18	14.42	2.31	3.04	.000***
Decl	0.82	0.22	3.63	0.38	1.26	.000***
Proc	0.07	0.18	0.41	−0.29	0.44	.684
Session	1.14	0.11	10.01	0.92	1.37	.000***
Decl:Session	−0.03	0.11	−0.28	−0.25	0.19	.780

Note: ***$p < .001$.

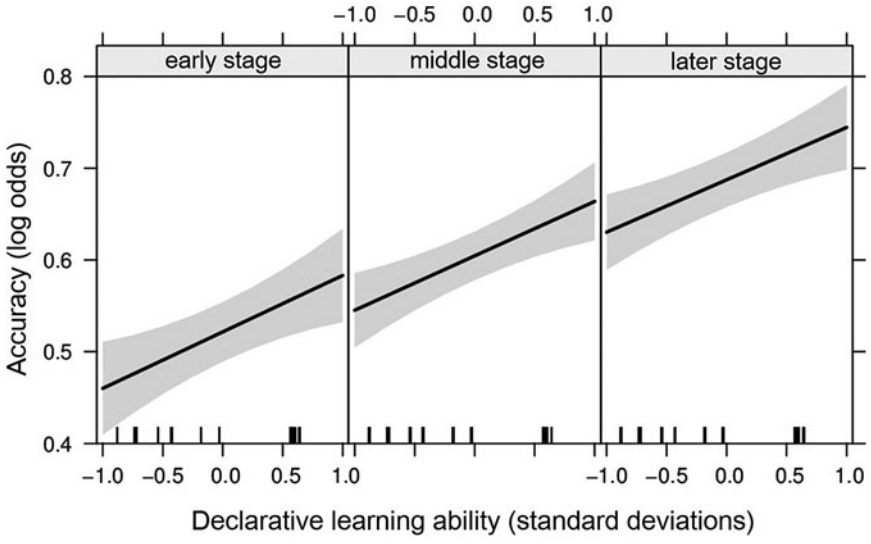


Fig. 1. Effect of declarative learning ability on accuracy in comprehension. Values on the x-axis represent standard deviations of the composite declarative learning ability score. The rugs along the x-axis of each panel represent the distribution of declarative learning ability values in the sample. Values on the y-axis represent the log odds of a correct response on a comprehension trial. The left, center, and right panels represent early, middle, and later stages of practice, respectively, and do not correspond directly to particular training blocks.

Table 4. Mixed-effects model of the effects of session, declarative learning ability and procedural learning ability on accuracy in production.

				95% CI		
Fixed effects	β	*SE*	*z*	lower	upper	*p*
(Intercept)	0.59	0.66	0.89	−0.70	1.89	.370
Decl	0.98	0.75	1.30	−0.50	2.46	.193
Proc	1.48	0.80	1.85	−0.09	3.06	.064^
Session	2.71	0.32	8.47	2.09	3.34	.000***
Proc:Session	−1.16	0.27	−4.27	−1.69	−0.63	.000***

Note: ^$p < .10$; ***$p < .001$.

accurate responses in later stages of practice compared to earlier stages. The plot in Figure 2 illustrates the effect of procedural learning ability at three subsequent stages corresponding to intervals representing early, middle, and later stages of practice.

RQ2

Descriptive statistics

The 20 comprehension practice trials from Block 1 (Session 1) were considered warm-up practice and excluded from analysis. The analyzed RT data included correct trials in the remaining comprehension blocks that were within ± 2SDs of the mean RT calculated for each of the four sessions. Overall, 6.2% of the correct responses in the comprehension data were outside of the ± 2SDs criterion and were not included in the analysis.

According to Segalowitz (2010) the CV is a reliable index of automatization if (a) both CV (the ratio between the individual standard deviation in RT responses at block level and the RT mean at block level) and RT significantly decrease across practice, and (b) CV and RT are significantly correlated. Table 5 presents a summary of mean CV and RT values averaged across participants for each session (plots of these values across all blocks are available

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

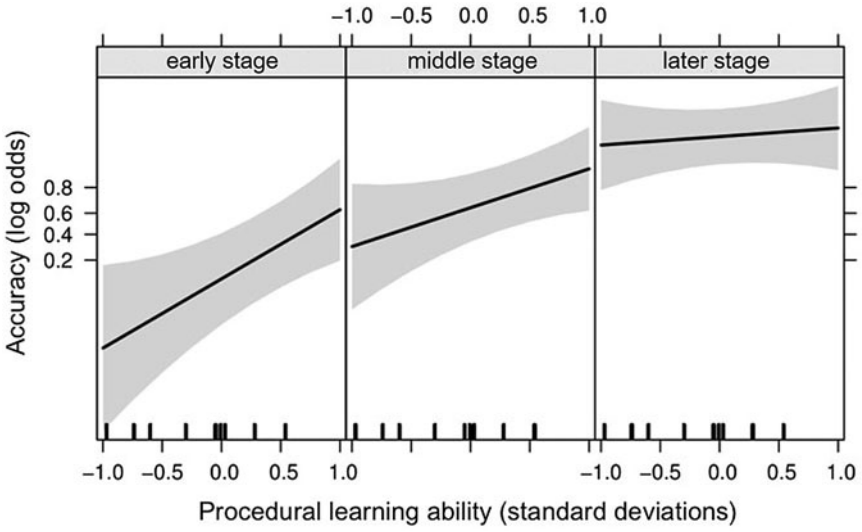


Fig. 2. Effect of procedural learning ability on accuracy in production. Values on the x-axis represent standard deviations of the composite procedural learning ability score. The rugs along the x-axis of each panel represent the distribution of procedural learning ability values in the sample. Values on the y-axis represent the log odds of a correct response on a production trial. The left, center, and right panels represent early, middle, and later stages of practice, respectively, and do not correspond directly to particular training blocks.

Table 5. Mean CV and RT (in milliseconds) across sessions (N = 14).

	S1	S2	S3	S4	Overall
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M(SD)*
CV	1.33 (0.3)	1.13 (0.3)	0.93 (0.2)	0.82 (0.2)	1.05 (0.2)
RT	5207 (1945)	2872 (1094)	1774 (620)	1465 (367)	2829 (748)

Table 6. Correlations between CV and learning ability across sessions (N = 14).

	S1	S2	S3	S4	Overall
Declarative	−.145	−.531	−.439	−.329	−.463
Procedural	−.504^	−.629^	−.563^	−.582^	−.743*

Note: ^p < .10; *p < .05. Bootstrapped; Holm-Bonferroni corrected.

as supplementary materials S2, Supplementary Materials). In regard to the first criterion, we find that both CV and RT decreased statistically significantly between Session 1 and Session 4 (for CV: t (13) = 5.23, p = .004, d = 1.7; for RT: t (13) = 6.83, p = .004, d = 2.7; bootstrapped). In regard to the second criterion, we calculated the CV and RT for each of the comprehension blocks included in the analysis, averaging across participants, and found that the correlation between CV and RT (r (33) = .746, p = .009; bootstrapped) was positive and statistically significant (see S2 for a plot). Thus, our data meet the criteria for CV to be interpreted as an index of automatization.

Next, we take a preliminary look at the relationship between CV and learning ability (Table 6). It is important to note that, as lower CV values indicate higher automatization, negative correlations between learning ability and CV indicate positive relationships of these variables with automatization. Over the sessions, we see a weak to medium relationship between CV and declarative learning ability and a medium to strong relationship between CV and procedural learning ability. The correlations relative to the overall CV mean scores reflect this pattern in that procedural learning ability, but not declarative learning ability, was found to significantly correlate with the coefficient of variation.

Data modeling

In order to directly address RQ2, data modeling was performed using mixed-effects models with the *lmer* function (*lme4* package, Bates et al., 2011) in the *R* environment (R Development Core Team, 2018), after a low risk of multicollinearity was ascertained (condition number = 1.45). The log-transformed CV ($\log_{10}$) was the dependent variable. The predictors were Decl and Proc (both standardized) and Session (continuous and centered). The derivation of the model followed the criteria illustrated earlier (cf. S1 for the model's syntax).

In the model output (Table 7), a negative β coefficient indicates a negative correlation between the predictor and the CV measure, hence a POSITIVE relationship between the predictor and automatization, as lower CV values indicate more automatization. Conversely, a positive β value indicates a NEGATIVE relationship between the predictor variable and automatization, as higher CV values indicate less automatization. Overall, the mixed-effects model explained 37% of the variance, compared to 11% in the corresponding model with no random effects.

A statistically significant, but small, effect of Session (R^2 = .11, p < .01) was observed indicating that session-dependent factors beyond learning ability contributed to increased automatization over time. Turning to the long-term memory predictors, the model showed that, overall, procedural learning ability had a statistically significant positive effect on automatization (p < .01) and accounted for about 30% of the variance (a medium effect), whilst declarative learning ability exerted a positive, small-sized effect (5% of the variance) but was not statistically significant.

The model also returned a statistically significant (p < .05) Decl by Proc by Session interaction. In discussing this result it is important to remember that the interaction, per se, does not imply any specific directionality or causality. As one of the

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

Table 7. Mixed-effects model of the effects of session, declarative learning ability and procedural learning ability on automatization.

				95% CI		
Fixed effects	β	*SE*	*t*	lower	upper	*p*
(Intercept)	−0.11	0.01	−7.31	−0.14	−0.08	.000***
Decl	−0.02	0.02	−0.98	−0.07	0.02	.377
Proc	−0.08	0.02	−4.31	−0.12	−0.04	.003**
Session	−0.04	0.01	−3.13	−0.06	−0.01	.007**
Decl:Proc	−0.04	0.02	−1.79	−0.08	0.00	.147
Decl:Session	0.02	0.02	1.13	−0.01	0.06	.282
Proc:Session	−0.02	0.01	−1.65	−0.04	0.00	.177
Decl:Proc:Session	−0.04	0.01	−2.55	−0.07	−0.01	.029*

Note: ^p < .10; *p < .05; **p < .01; ***p < .001.

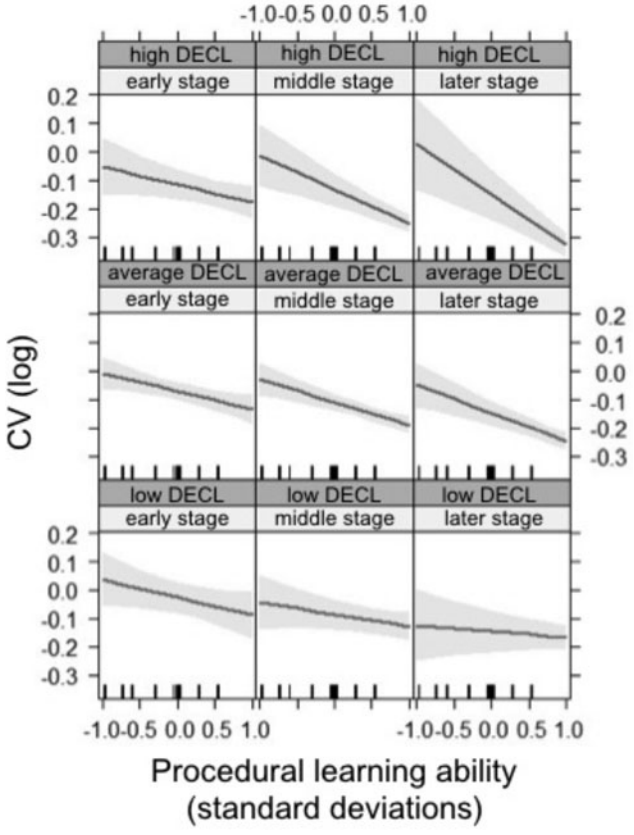


Fig. 3. Effect of the DECL by PROC by SESSION interaction on automatization. Values on the x-axis represent standard deviations of the composite procedural learning ability score. The rugs along the x-axis of each panel represent the distribution of procedural learning ability values in the sample. Values on the y-axis represent the log of the CV index. Panels from left to right represent the effect of procedural learning ability for early, middle and later stages of practice for a constant level of declarative learning ability. Panels from bottom to top represent the effect of procedural learning ability for increasing levels of declarative learning ability at a given stage of practice. Note that levels of declarative learning ability and stages of practice in the figure represent intervals in the values of the respective continuous variables generated for the purpose of illustration and, thus, should not be interpreted as factor levels.

possible illustrations of the interaction, we plot the effect of procedural learning ability from the model for different levels of declarative learning ability across practice (Figure 3).

Reading the plot from left to right (and keeping the stage in practice constant), we note that in the early stages of practice ('early stage') declarative and procedural learning ability do not appear to interact: that is, the slope of procedural learning ability is virtually the same regardless of the level of declarative learning ability. The effect of the interaction emerges in the middle stage of training ('middle stage'), and, even more clearly, later in training ('later stage'). At those stages, declarative and procedural learning ability do appear to interact in that the slope of procedural learning ability becomes steeper and more negative for higher levels of declarative learning ability. Thus, later in practice, better procedural learning ability is associated with more automatization for learners with higher declarative learning ability.

The same interaction can also be viewed in another manner: reading the plot from top to bottom (and keeping the DECL level constant), we note that, for average and above-average values of declarative learning ability ('average DECL' and 'high DECL'), higher procedural learning ability is associated with steeper, more negative slopes representing better automatization over the course of practice. For below-average levels of declarative learning ability ('low DECL'), the procedural memory effect seems to flatten out over practice, suggesting that automatization becomes markedly worse over the course of practice as procedural learning ability increases.

Overall, the plot of the three-way interaction seems to indicate at least two facts: (a) that the interaction between long-term memory abilities does not emerge immediately and (b) that the effect of procedural learning ability on automatization varies differently over time for learners with different levels of declarative learning ability. As illustrated in Figure 3, higher declarative learning ability increasingly supports the effect of procedural learning ability on automatization. However, lower declarative learning ability is detrimental for the effect of procedural learning ability on automatization as practice progresses.

Note that all data and analyses for each of the above research questions are available on the Open Science Foundation page – osf.oi/uzw6r.

Discussion

The first research question asked TO WHAT EXTENT DECLARATIVE AND PROCEDURAL LEARNING ABILITY PREDICTED ACCURACY IN COMPREHENSION AND PRODUCTION IN L2 PRACTICE, AND WHETHER THESE EFFECTS VARIED

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

ACROSS PRACTICE. For comprehension practice, the mixed-effects model analysis revealed a positive, medium, statistically significant relationship between declarative learning ability and accuracy; whereas, for procedural learning ability, no statistically significant relationship with accuracy was detected. We also found that comprehension accuracy improved over the sessions, but this effect did not interact with either declarative or procedural learning ability, indicating that their relationships with accuracy did not vary significantly across practice. A strong role for declarative learning ability in predicting accuracy during practice is consistent with the previously discussed findings in Pili-Moss (2018, Study 2), where learners engaged in a total of six blocks of 20 comprehension practice trials.

Our finding that declarative learning ability was related to comprehension accuracy early in practice is consistent with the results of the meta-analysis in Hamrick et al. (2018), and in particular with the results in Morgan-Short et al. (2014), the study from which our data were obtained. However, discrepancies with Morgan-Short et al. (2014), and more generally with the results reported in Hamrick et al.'s meta-analysis, emerge with regard to the findings at later stages of practice in at least two respects. First, the GJT findings in Morgan-Short et al. indicated that the effect of declarative learning ability became nonsignificant after the end of practice, whilst in our study it slightly decreased across practice, but not significantly. Second, Morgan-Short et al. found that procedural learning ability predicted accuracy on the GJT after the end of practice, whilst no significant effect of procedural learning ability emerged in comprehension practice in the present study.

Considering that the present study analyzes accuracy taken from the same participants in the same experiment, the question emerges of why, contrary to results for accuracy on the GJT (Morgan-Short et al., 2014), the declarative learning ability effect for accuracy on practice was maintained even at later stages and the procedural learning ability effect was not evidenced at any point. One possibility is simply that the GJT was administered at only two time points, after certain amounts of practice had been completed, whereas practice was continuous. Another possibility is that the type of task used to measure accuracy had an effect on the engagement of declarative and procedural learning ability during practice, a possibility already envisaged in Morgan-Short et al. (2014, p. 69). For example, even though participants did not receive instructions to search for rules, they were likely to apply hypothesis testing to work out strategies to improve their score, which reflected the accuracy of their responses during practice. Evidence that rule-based tasks, which can be learned via explicit hypothesis testing, activate neural areas that implicate declarative memory has been discussed in studies of human category learning (e.g., Ashby & Crossley, 2012, for a review). Also, it is possible that declarative memory was more engaged during practice due to the fact that participants had to process/retrieve arbitrary aural-visual associations (Henke, 2010). It is known that the integration of multiple cues in a task, particularly if the cues are visual-spatial, specifically engages declarative memory (Packard & Goodman, 2013; Ullman, 2016).

By contrast, the GJT in Morgan-Short et al. (2014) only required learners to evaluate aural stimuli in a situation where, due to lack of visual-spatial associations in the stimuli, declarative processing was arguably less compelling, with consequent greater reliance on procedural processing. Overall, we conclude that the asymmetry between L2 practice and GJT in the relationship with long-term memory abilities may point towards an enhanced role of declarative learning ability that may be due to the processing requirements of the gaming task.

Now turning to production practice, the mixed-effects model analysis did not detect a statistically significant relationship between production accuracy and either declarative or procedural learning ability. However, an effect of procedural learning ability was stronger at early stages of practice and significantly decreased as practice progressed. These results do not seem fully consistent with the results from Morgan-Short et al. (2014), where a relationship between procedural learning ability and accuracy on a GJT was detected at the end of practice, but not after the first session of practice. We can speculate that the difference in this pattern of results, again, might emerge because of the type of task that learners were engaged in during practice as opposed to during the GJT, although exactly why this should be the case remains unclear.

A related question is why the effect of procedural learning ability declined as training progressed. We offer two speculative reasons for this finding. One possibility is that, unlike participants with low procedural learning ability, participants with high levels of procedural learning ability may have been able to benefit from lower amounts of input early on in practice. With increasing amounts of input, differences in attainment between low and high levels of procedural learning ability might have leveled off. A second possibility that might also be considered involves the relationship between comprehension and production in L2 development (cf. De Jong, 2005; DeKeyser & Sokalski, 2001; Izumi, 2003; Ellis, 2005), and specifically the hypothesis that input processing in comprehension may feed into processing in production, in particular when the process involves declarative knowledge. Assuming that the initial effect of procedural learning ability reflects a very early stage in L2 processing at which comprehension (strongly driven by declarative memory) does not yet feed into production, the relationship between comprehension and production could strengthen later in practice, and processing during production become less reliant on procedural learning ability as a consequence.

The second research question asked TO WHAT EXTENT DECLARATIVE AND PROCEDURAL LEARNING ABILITY PREDICTED AUTOMATIZATION IN LANGUAGE COMPREHENSION ACROSS PRACTICE, i.e., to what extent they predicted negative values of the coefficient of variation. First of all, the analysis showed that the pattern of CV scores across practice was compatible with L2 automatization in comprehension, i.e., both CV and RT significantly decreased across practice, and there was a significant correlation between them. This supports findings of previous studies using the CV to investigate automatization of L2 syntax (e.g., Lim and Godfroid, 2015; Ma et al., 2017).

With regard to the cognitive variables of interest, the analysis showed that procedural learning ability had a positive, medium, significant effect on automatization, whereas declarative learning ability had a positive, small effect that was not statistically significant. However, these effects were conditional to a significant three-way interaction with session that indicated that automatization in comprehension benefitted from an interaction between declarative and procedural learning ability during processing, and increasingly so later in practice. Inspection of the plot in Figure 3 showed that the interaction did not emerge immediately, but only after the participants had had some initial practice with the language. Additionally, the interaction indicated an association between higher procedural learning ability and greater automatization that became stronger with practice for learners with higher declarative learning ability. For learners with lower

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

levels of declarative learning ability, the interaction indicated that higher procedural learning ability was comparatively not as beneficial for automatization at later stages of practice.

Overall, these findings support the close link between behavioral measures of procedural memory and L2 automatization, a relationship that has been often implied in the literature but for which behavioral evidence has only recently started to emerge. Recently, Suzuki (2017) found that procedural memory correlated with RT reduction (an element of automatization), although no relationship between procedural memory and automatization was evidenced. By contrast, the present study found a significant relationship between the CV and procedural learning ability as well as a significant interaction between declarative and procedural learning ability that varied across practice. It is possible that the discrepancy in results depends on methodological differences between the two studies, such as the fact that, unlike ours, Suzuki's study administered explicit L2 instruction, deployed a single task (the TOL) to measure procedural memory, and analyzed production instead of comprehension.

The results of the present study are also compatible with the predictions that some current cognitive approaches to L2 learning would make for the engagement of declarative and procedural resources in L2 learning and processing (e.g., DeKeyser, 2015; Paradis, 2009; Ullman, 2015). In terms of the effects of declarative and procedural memory for L2 learning, the results relative to the analysis of accuracy in comprehension are in line with neurocognitive models that predict a significant engagement of declarative memory in the initial stages of L2 learning (Paradis, 2009; Ullman, 2004, 2015, 2016). This effect is due to the specific capability of the declarative memory system to learn efficiently in conditions of limited input. We have argued that the fact that the strength of this effect appears to mitigate to a lesser extent during practice, compared to when L2 proficiency is measured with a GJT, may indicate that an additional effect of task is at play that further biases processing towards the declarative modality.

Unlike Paradis (2009), Ullman's DP model would also be compatible with the significant role of procedural learning ability for automatization found in the present study. This is because Ullman's DP model would not exclude a role for procedural memory in conditions of relatively limited exposure to a second language such as the ones provided in our experiment. Both declarative and procedural memory may be contributing to language development at any stage with the relative strength of their effect varying over time.

A further aspect that is very generally compatible with Ullman's model is the finding of a significant interaction between declarative and procedural learning ability during processing. Ullman discusses that declarative and procedural memory may cooperate or compete with each other, based on evidence from human and animal studies that has accumulated in neuropsychology and neuroscience in the last fifty years (Packard & Goodman, 2013). The finding of an interaction in our results (Figure 3) suggests that the relationship between the two memory systems may depend, among other possible factors, on individual strengths within the systems. We see cooperation when individuals have high declarative learning ability, but competition when individuals' declarative learning ability is below average. Compatible with a cooperative interaction interpretation, Morgan-Short et al. (2015) also found that engagement of procedural memory neural substrates in individuals with high declarative memory enhanced L2 proficiency at initial stages of practice.

Further, these results are largely compatible with other theoretical models that posit a supporting role of declarative knowledge in the establishment of proceduralized L2 knowledge (e.g., DeKeyser, 2015; Ellis, 2005). Specifically, in line with the predictions of DeKeyser (2007, 2015), automatization in comprehension is significantly related to procedural processing, and increasingly so as practice progresses, whereas the effect of declarative learning ability declines across practice. Furthermore, the overall positive effect for automatization of the interaction between declarative and procedural learning ability indicates that (high levels of) declarative learning ability reinforce the capacity of procedural learning ability to predict automatization (and vice versa). Although the interaction per se does not indicate the direction of the effect, the results are compatible with the interpretation that, in the early stages of automatization, declarative learning ability may perform a supporting/ancillary function with respect to procedural learning ability, which remains the main engine of the process.

Overall, the results from the present analysis of L2 practice are largely compatible with the predictions recent cognitive models have made with regard to the engagement of declarative and procedural memory/knowledge in L2 learning and processing and their interaction. This is particularly the case for the analysis of L2 accuracy in comprehension and for automatization in comprehension.

Limitations of the study and further research

The study has a number of limitations that should be addressed by further research. First, in the analysis of both accuracy and automatization, the effects of comprehension on production (and vice-versa) were not controlled. Specifically, participants were administered comprehension as well as production practice blocks, and it is possible that L2 processing in one modality may have affected L2 processing and attainment in the other. Future research could seek to control these effects: for example, by adopting experimental designs where type of practice is a between-group variable.

Secondly, although the large number of trial items ensured the viability of the inferential analysis using mixed-effects models, it is of paramount importance that the effects of long-term memory abilities during practice are investigated more extensively in studies with a larger number of participants.

Further, the analysis of automatization in the present study was partial because it only examined comprehension practice. Further research could investigate how the development of automatization varies in comprehension and production overall, as well as specifically look at the effects of declarative and procedural learning ability in the two modalities. A further important aim in this line of research should be to design studies that elucidate whether and how a wide set of factors – including, for example input complexity and the extent to which L2 knowledge is explicit – modulate the effect of long-term memory in automatization. Additionally, the analysis of automatization in the present study deployed the CV index as the outcome measure. It remains to be shown whether results would be confirmed if alternative measures of automatization were used: for example, a measure based on the fit of individual latency data to a power function. Similarly, it will be important for researchers to show that the patterns of results are robust over different measures that are valid measures of declarative and procedural memory (for preliminary work on this issue, see Buffington & Morgan-Short, 2019).

A further development of interest would be to include additional cognitive variables in the study of both L2 accuracy and automatization. For instance, alongside declarative and procedural

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

learning ability, one could investigate the role of working memory as a main effect, as well as a potential moderating effect in an interaction. Specifically, since working memory is known to support declarative processing, and a significant role of declarative learning ability has been found for both L2 accuracy and L2 automatization, a study with a design similar to the present one could explore to what extent working memory modulates declarative learning ability. Finally, future studies could investigate the role of long-term memory individual differences for L2 accuracy and automatization across a wider range of linguistic structures and, possibly, different age groups.

Conclusions

This study offered an exploratory analysis of the effects of declarative and procedural learning ability on L2 accuracy and automatization during language practice over the course of two weeks. The study found distinct patterns in the effects of the two learning abilities in comprehension accuracy, production accuracy, and comprehension automatization. Declarative learning ability emerged as the main predictor of accuracy in comprehension, an effect that did not significantly change across practice. However, neither learning ability was a significant predictor of accuracy in production, although we found that procedural learning ability predicted production accuracy more at early stages and significantly less later in practice. This pattern of results differs from what had been found in the same set of learners for performance on GJTs administered after one session of practice and after the end of practice. We have suggested that, at least for comprehension accuracy, the discrepancy in the findings may be largely due to the type of task.

By contrast, procedural learning ability was a main predictor of automatization in comprehension, a finding that, to the best of our knowledge, had not yet been reported in a behavioral experiment. A further predictor that on average supported automatization was an interaction between declarative and procedural learning ability. Overall, these results support predictions of the DP model with regard to the prominence of declarative processing early in practice, as well as with regard to the possibility of cooperative interactions between declarative and procedural memory in L2 development (Ullman, 2004, 2015, 2016). Likewise, the study supports key predictions Skill Acquisition Theory makes for the proceduralization of L2 skills during practice (DeKeyser, 2007, 2015), including the finding that procedural learning ability was a significant predictor of automatization and that declarative learning ability appeared to support automatization in its early stages.

Overall, extending previous research, the present study found that long-term memory plays a pivotal role in accounting for the development of L2 accuracy and automatization during practice. By examining the effect of learning abilities during L2 practice we may have further insight into the role the declarative and procedural memory systems play in the learning process.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728919000543

Acknowledgments. We thank Michael Ullman for discussion at initial stages of the development of the study, the audience of the Sixth Implicit Learning Seminar (ELTE, Budapest, May 2017) for constructive feedback and three anonymous reviewers for their insightful suggestions. All errors remain our own.

References

Akamatsu N (2008) The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics* **29**, 175–193. doi:10.1017/S0142716408080089

Anderson JR (1993) *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson JR (2007) *How can the human mind occur in the physical universe?* New York: Oxford University Press. doi:10.1093/acprof:oso/9780195324259.001.0001

Antoniou M, Ettlinger M and Wong PCM (2016) Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLOS One* **11**, e0158812. doi:10.1371/journal.pone.0158812

Ashby FG and Crossley MJ (2012) Automaticity and multiple memory systems. *Wiley Interdisciplinary Reviews: Cognitive Science* **3**, 363–376. doi:10.1002/wcs.1172

Bates D, Machler M and Bolker B (2011) lme4: Linear mixed-effects models using s4 classes. http://cran.R-project.org/package=lme4.

Barr D.J., Levy R, Scheepers C and Tily HJ (2013) Random effect structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Brill-Schuetz KA and Morgan-Short K (2014) The role of procedural memory in adult second language acquisition. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 260–265.

Buffington J & Morgan-Short K (2019) Declarative and Procedural Memory as Individual Differences in Second Language Aptitude. In Wen Z, Skehan P, Biedroń A, Li & S Sparks R (eds), *Language aptitude: Advancing theory, testing, research and practice*. New York: Routledge, pp. 215–237.

Cabeza R and Moscovitch M (2013) Memory systems, processing modes, and components: Functional neuroimaging evidence. *Perspectives on Psychological Science* **8**, 49–55. doi:10.1177/1745691612469033

Carpenter HS (2008) *A behavioral and electrophysiological investigation of different aptitudes for L2 grammar in learners equated for proficiency level*. Ph.D. dissertation. Georgetown University.

Carroll JB and Sapon SM (1959) *Modern Language Aptitude Test*. New York: The Psychological Corporation/Harcourt Brace Jovanovich.

De Jong N (2005) Can second language learning be learned through listening?: An experimental study. *Studies in Second Language Acquisition* **27**, 205–234. doi:10.1017/S0272263105050114

DeKeyser RM (1995) Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition* **17**, 379–410.

DeKeyser RM (1997) Beyond explicit rule learning. *Studies in Second Language Acquisition* **19**, 195–221.

DeKeyser R (2007) Skill acquisition theory. In VanPatten B & Williams J (eds), *Theories in second language acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum, pp. 97–113.

DeKeyser RM (2015) Skill acquisition theory. In VanPatten B and Williams J (eds), *Theories in second language acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 94–112.

DeKeyser RM and Sokalski KJ (2001) The differential role of comprehension and production practice. *Language Learning* **51**, 81–112.

Eichenbaum H (2008) *Learning & Memory*. New York: W.W. Norton & Company.

Eichenbaum H (2011) *The cognitive neuroscience of memory: An introduction*. New York: Oxford University Press.

Ellis NC (2005) At the interface: Dynamic interaction of explicit and implicit knowledge. *Studies in Second Language Acquisition* **27**, 305–352.

Ettlinger M, Bradlow AR and Wong PCM (2014) Variability in the learning of complex morphophonology. *Applied Psycholinguistics* **35**, 807–831. doi:10.1017/S0142716412000586

Faraway JJ (2016) *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Boca Raton: CRC Press.

Faretta-Stutenberg M and Morgan-Short K (2018) The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research* **34**, 67–101. doi:10.1177/0267658316684903

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press

Ferman S, Olshtain E, Schechtman E and Karni A (2009) The acquisition of a linguistic skill by adults: Procedural and declarative memory interact in the learning of an artificial morphological rule. *Journal of Neurolinguistics* **22**, 384–412. doi:10.1016/j.jneuroling.2008.12.002

Foerde K, Knowlton BJ and Poldrack R (2006) Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences* **103**, 11778–11783.

Friederici AD, Steinhauer K and Pfeifer E (2002) Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences* **99**, 529–534.

Hamrick P (2015) Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences* **44**, 9–15. doi:10.1016/j.lindif.2015.10.003

Hamrick P, Lum JAG and Ullman MT (2018) Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences* **115**, 1487–1492. doi:org/10.1073/pnas.1713975115

Henke K (2010) A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience* **11**, 523–532. doi:10.1038/nrn2850

Hulstijn JH, Van Gelderen A and Schoonen R (2009) Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics* **30**, 555-–82. doi:10.1017/S0142716409990014

Izumi S (2003) Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the Output Hypothesis. *Applied Linguistics* **24**, 168–196. doi:10.1093/applin/24.2.168

Kaller CP, Rahm B, Köstering L and Unterrainer JM (2011) Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research* **216**, 1–8.

Kaller CP, Unterrainer JM and Stahl C (2012) Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment* **24**, 46–53.

Kaufman AS and Kaufman NL (2004) *The Kaufman Brief Intelligence Test, Adult Version, Second Edition (K-BIT-2)* (2nd edn.) Circle Pines, MN: American Guidance Service.

Lim H and Godfroid A (2015) Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics* **36**, 1247–1282. doi:10.1017/S0142716414000137

Ma D, Yu X and Zhang H (2017) Word-Level and Sentence-Level Automaticity in English as a Foreign Language (EFL) Learners: A Comparative Study. *Journal of Psycholinguistic Research* **46**, 1471–1483. doi:10.1007/s10936-017-9509-8

Morgan-Short K (2007) *A neurolinguistic investigation of late- learned second language knowledge: The effects of explicit and implicit conditions*. Ph.D. dissertation, Georgetown University.

Morgan-Short K, Deng Z, Brill-Schuetz KA, Faretta-Stutenberg M, Wong P and Wong FCK (2015) A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition* **37**, 383–419.

Morgan-Short K, Faretta-Stutenberg M, Brill-Schuetz K, Carpenter H and Wong PCM (2014) Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition* **17**, 56–72. doi:10.1017/S1366728912000715

Morgan-Short K, Finger I, Grey S and Ullman MT (2012) Second language processing shows increased native-like neural responses after months of no exposure. *PLOS One* **7**, e32974. doi:10.1371/journal.pone.0032974

Morgan-Short K, Sanz C, Steinhauer K and Ullman MT (2010) Second language acquisition of gender agreement in explicit and implicit training conditions: An event- related potential study. *Language Learning* **60**, 154–193.

Morgan-Short K, Steinhauer K, Sanz C and Ullman MT (2012) Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience* **24**, 933–947.

Norris JM and Ortega L (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* **50**, 417–528.

Packard MG and Goodman J (2013) Factors that influence the relative use of multiple memory systems. *Hippocampus* **23**, 1044–1052. doi:10.1002/hipo.22178

Paradis M (2009) *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins Publishing Company.

Phillips NA, Segalowitz N, O'Brien I and Yamasaki N (2004) Semantic priming in a first and second language: evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics* **17**, 237–262. doi:10.1016/S0911-6044(03)00055-1

Pili-Moss D (2018) *The earliest stages of second language learning: A behavioral investigation of long-term memory and age*. Ph.D. dissertation, Lancaster University. www.research.lancs.ac.uk/portal/files/260180551/PhD_Diana_Pili_Moss_registry.pdf (retrieved June 1st, 2019)

Plonsky L and Ghanbar H (2018) Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*.

Plonsky L and Oswald FL (2014) How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning* **64**, 878–912.

R Development Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Segalowitz N (2003) Automaticity and second language learning. In Doughty C & Long M (eds), *The handbook of second language acquisition*. Oxford: Blackwell, pp. 382–408. doi:10.1002/9780470756492

Segalowitz N (2010) *Cognitive bases of second language fluency*. Routledge: New York/London.

Segalowitz NS (2013) Automaticity. In Robinson P (ed.), *The Routledge encyclopedia of second language acquisition*. London: Routledge, pp. 53–57.

Segalowitz NS and Segalowitz SJ (1993) Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics* **14**, 369–369.

Segalowitz SJ, Segalowitz NS and Wood AG (1998) Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics* **19**, 53–67. doi:10.1017/S0142716400010572

Segalowitz N, Trofimovich P, Gatbonton E and Sokolovskaya A (2008) Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon* **3**, 47–71.

Squire LR (2004) Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* **82**, 171–177.

Squire LR and Dede A. J. O. (2015) *Conscious and unconscious memory systems. Cold Spring Harbor Perspectives in Biology* **7**(a021667) doi:10.1101/cshperspect.a021667

Squire LR and Wixted JT (2011) The cognitive neuroscience of human memory since HM. *Annual Review of Neuroscience* **34**. doi:10.1146/annurev-neuro-061010- 113720

Suzuki Y (2017) The role of procedural learning ability in automatization of L2 morphology under different learning schedules [First view online publication]. *Studies in Second Language Acquisition*, doi:10.1017/S0272263117000249. Published online by Cambridge University Press, August 10, 2017.

Suzuki Y and Sunada M (2018) Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. *Bilingualism: Language and Cognition* **21**, 32–46. doi:10.1017/S1366728916000857.

Trahan DE and Larrabee GJ (1988) *Continuous Visual Memory Test*. Odessa, FL: Assessment Resources.

Ullman MT (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* **92**, 231–270. doi:10.1016/j.cognition.2003.10.008

Ullman MT (2015) The declarative/procedural model: A neurobiologically-motivated theory of first and second language. In VanPatten B and Williams J (eds), *Theories of second language acquisition: An introduction*. Mahwah: NJ: Lawrence Erlbaum Associates, pp. 135–158.

Ullman MT (2016) The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In Hickok G and Small SA (eds), *The neurobiology of language*. Elsevier, pp. 953–968. doi.org/10.1016/B978-0-12-407794-2.00092-4

Unterrainer JM, Rahm B, Leonhart R, Ruff CC and Halsband U (2003) The Tower of London: The impact of instructions, cueing, and learning on planning abilities.

https://doi.org/10.1017/S1366728919000543 Published online by Cambridge University Press